DOI 10.52209/1609-1825_2023_1_365 ӘОЖ 004.912

Мәтіндерді функционалдық стилі бойынша классификациялаудың оңтайлы моделі

[1]*КАЙБАСОВА Динара Женисбековна, PhD, доцент м.а., dindgin@mail.ru,
[1]ОЛЕЙНИКОВА Алла Васильевна, магистр, аға оқытушы, alla_ole@mail.ru,
[1]МАХАНОВА Балжан Мақсатқызы, магистрант, balzhan.makhanova@gmail.com,
[1]«Әбілқас Сағынов атындағы Қарағанды техникалық университеті» КеАҚ, Қазақстан, Қарағанды, Н. Назарбаев даңғылы, 56,
*автор-корреспондент.

Аңдатпа. Бүгінгі таңда табиғи тілді өңдеу деректерге қол жетімділіктің айтарлықтай жақсаруы мен есептеу қуаттылығының артуының арқасында дамып келеді, атап айтқанда, денсаулық сақтау, бұқаралық ақпарат құралдары, қаржы және адам ресурстары сияқты салаларда маңызды нәтижелерге қол жеткізуге мүмкіндік береді. Ұсынылған мақаланың мақсаты мәтіндердің стилін анықтауда мәтіндерді интеллектуалдық талдау әдістерін зерттеу және оңтайлы моделін құру. Мәтіндер стилін анықтауда қолданылатын мәтіндік деректер көздері қарастырылды. Мәтіндер корпусы толық өңделіп, қажетті әдістерді қолдана отырып эксперименттер жасалды. Мәтіндерді классификациялауда К-жақын көршілер және BERT классификация модельдері қолданылды.

Кілт сөздер: мәтіндерді талдау, мәтін стилі, функциональдық стиль, классификация, К-жақын көршілер, BERT, табиғи тілдер, мәтінді зияткерлі талдау, көркем әдебиет стилі, ғылыми стиль, публицистикалық стиль.

Кіріспе

Табиғи тілдерді өңдеудің дәстүрлі әдістерінің компьютерлік лингвистикадан айырмашылығы, біріншісінің лингвистика жалпы зерттейтін барлық нәрсені модельдеуге, ал екіншісі табиғи тілдерді сипаттауға қабілетті математикалық модельдерді құруға бағытталғандығында [1]. Компьютерлік лингвистиканың негізгі міндеті модельдерді құру және сәйкес алгоритмдер мен мәтіндерді автоматты түрде өңдеуге арналған бағдарламалар ретінде тұжырымдалуы мүмкін.

Мәтінді компьютерлік талдау – бұл үлкен көлемдегі мәтіндерді талдау және олардан ақпарат алу үшін есептеу құралдарын қолдану әдісі. Бұл әдіс жеке мәтіндерден бастап үлкен (мәтіндік) мәліметтерге дейін сандық мәтіндерді талдау үшін компьютерлер мен бағдарламалық жасақтаманың күшін пайдаланатын сандық құралдар мен сандық әдістер жиынтығының жалпы термині. Табиғи тілдегі мәтінді компьютерлік талдау соңғы жылдары көптеген бағыттарда белсенді дамып келеді. Қазіргі кездегі есептеу қуаты іздеу, жіктеу, кластерлік талдау, мәліметтердегі жасырын заңдылықтарды анықтау және т.б. мәселелерін тиімді шешуге ықпал ететін құжаттардың үлкен массивтерін өңдеуге арналған математикалық әдістердің кең класын қолдануға мүмкіндік береді.

Автоматты талдауда қажетті құрылымдарды таңдау үшін бірнеше дәйекті кезеңдерден өту керек:

1) бастапқы мәтін
2) алдын-ала талдау
3) морфологиялық талдау
4) таяз талдау
5) терең талдау
6) таяз семантикалық талдау
7) терең мағыналық талдау
8) прагматикалық талдау
9) мәтіндік құрылымдарды сәйкестендіру [2].

Қазіргі уақытта мәтіннің қандай функционалды стильге жататынын анықтау қажеттілігі кең тараған. Сондай-ақ, бірінші кезекте оқу немесе ғылыми мәтіндерді автоматты түрде өңдеу және сапасын бағалау міндетін атап өткен жөн, мұнда жарияланған стильдің нақты қолданылатын стильге сәйкестігін бағалау маңызды болып табылады. Бұндай бағыттар жазбаларды басқару, сондай-ақ, күнделікті әрекеттерге қатысты құжаттарды табу және талдау үшін мәтінді өңдеуді пайдалана алады. Жалпы бұл мәселе екі тәсіл негізінде шешіледі. Бірінші, орындауға оңай, ережелерді қолданатын адамның мәтінді талдауға қатысуын және мәтіннің белгілі бір стильге жататындығы туралы қорытындыны білдіреді. Жүзеге асырудың қарапайымдылығымен қатар бұл тәсілдің екі елеулі кемшілігі бар: мәтіндерді оқуға, талдауға және негіздеуге кететін уақыттың қажеттілігі; сонымен қатар стильдік классификация сапасының

тұлғаның тілдік дайындығына тәуелділігі.

Екінші тәсіл стильдік белгілердің мәндерін есептеу және кейбір математикалық аппараттар негізінде классификаторды синтездеу үшін компьютерлік мәтінді өңдеуді қолдануды қамтиды. Көптеген қолданбалар үшін бұл ыңғайлы, бірақ ешбір бағдарлама лингвистті алмастыра алмайтынын түсіну керек.

Мәтіндердің стилін анықтау үшін деректерді өңдеуге қажетті құрылымданбаған деректерді зерттеу көптеген жұмыстарда қарастырылған. Мысалға, [3] жұмыстың авторлары LSTM (ұзақ қысқа мерзімді жады) модульдері бар терең нейрондық желіні пайдалана отырып, мәтіндік құжаттарды жіктеу әдісін ұсынады. Бұл жұмыста жіктелетін құжаттарды білдіретін ерекшелік векторларын құрудың әртүрлі тәсілдері сыналған: құжаттарға енгізілген сөздер тізбегі ретінде құрастырылған ерекшелік векторлары пайдаланылды. Балама ретінде сөздер алдымен word2vec құралының көмегімен векторлық көріністерге түрлендірілді және осы векторлық көріністердің тізбегі құжат мүмкіндіктері ретінде пайдаланылды. Құжаттарды тақырыптық жіктеу міндетіне бұл тәсілдің қолданылуы 7 тақырыптық санатты білдіретін Уикипедия мақалаларының жинағы арқылы бағаланады. Тәжірибелер word2vec векторларында кодталған сөздер тізбегі ретінде ұсынылған құжаттармен LSTM желісіне негізделген тәсіл сөз жиілігі мүмкіндігінің векторлары ретінде ұсынылған құжаттармен стандартты сөздер жиыны тәсілінен асып түсетінін көрсеткен.

Келесі жұмыста [4] ұсынған модель сөз көріністерін үйрену кезінде максималды мүмкін болатын контекстік ақпаратты алу үшін қайталанатын құрылымды пайдаланған, бұл дәстүрлі терезе негізіндегі нейрондық желілермен салыстырғанда айтарлықтай азырақ шуды енгізе алады. Сондай-ақ мәтіндердің негізгі компоненттерін түсіру үшін мәтінді жіктеуде қандай сөздер негізгі рөл атқаратынын автоматты түрде анықтайтын максималды біріктіру қабаты қолданылады. Эксперименттік нәтижелер ұсынылған әдіс бірнеше деректер жиынына, әсіресе құжат деңгейіндегі деректер жиынына арналған заманауи әдістерден асып түсетінін көрсетті. Модель қайталанатын құрылымы бар контекстік ақпаратты түсіреді және конволюционды нейрондық желіні пайдаланып мәтіннің көрінісін жасайды. Эксперимент төрт түрлі мәтінді жіктеу деректер жиынын қолдана отырып, модельдің CNN және RecursiveNN-ден асып түсетінін көрсеткен [4].

BPNN (кері таралу нейрондық желілері) және терминді салмақтау әдісі ретінде қатыстылық коэффициенті (rf) көмегімен мәтінді жіктеу нәтижесін көрсетеді. Бұл мақала мәтінді жіктеу тапсырмасы үшін тиімді BPNN нейрондық желі алгоритмі бойынша жұмысты ұсынады, 3 санатқа жататын 150 мәтіндік құжат, яғни информатика, спорт және медицина Newsgroup20 тоқтату сөзді жою арқылы алдын ала өңделген, Портер алгоритмі, содан кейін екі басқарылатын өңдеу арқылы өңдеу. (tf) және (tf.rf) терминдерін салмақтау әдістері, эксперимент нәтижесі (tf.rf) (tf) қарағанда жақсы орындайтынын көрсетеді. Өнімділік F-өлшемі терминімен өлшенеді. Сәйкестік коэффициентін қолданбай, f шамасының мәні 0,72, tf .rf пайдаланғанда 0,92 нәтиже берген [5].

Жұмыста ұзақ мерзімді қысқа мерзімді жады (LSTM) архитектурасын пайдалана отырып, қайталанатын нейрондық желіні терең оқыту әдістерінің бірін қолдану арқылы эксперимент жүргізілді. Бұл зерттеуде модель LSTM және Global Vector (GloVe) көмегімен 300 өлшемді сөзді ендіру функцияларын пайдаланатын сынақ және қате эксперименттеріне негізделген. Параметрлерді баптау және ұсынылған сегіз LSTM-ді ауқымды деректер жинағымен салыстыру арқылы GloVe мүмкіндіктері бар LSTM мәтінді жіктеуде жақсы өнімділікті қамтамасыз ете алатыны көрсетілген. Нәтижелер GloVe бар LSTM көмегімен мәтінді жіктеу алтыншы модельде ең жоғары дәлдікті 95,17, орташа дәлдік, еске түсіру және F1 ұпайы 95 беретінін көрсетеді. Сонымен қатар, GloVe функциясы бар LSTM жақсы нәтижелерге жақын графикалық нәтижелер береді [6].

Материалдар мен әдістер

Мәтіннің функционалдық стилін автоматты түрде анықтау үшін нейрондық желілік технологияларды қолдануын зерттеу негізінде алға қойылған мақсатқа жету үшін біршама есептерді шешу қажеттігі туындайды. Атап айтқанда, стилистикалық үлгілерді талдау, нейрондық желіні жаттықтыру және сынау үшін әртүрлі стильдегі мәтіндердің белгілі бір жинағын теру қажет. Бұл тапсырманың қайнар көзі Орыс тілінің ұлттық корпусы болды. Сонда мәтіннің функциональдық стилін анықтау үшін келесі жиындар қарастырылды:

$$D = \{d_1, d_2, ..., d_n\},$$
$$Y = \{y_1(d_1), y_2(d_2), ..., y_n(d_n)\}.$$

мұндағы D – мәтіндер жинағы;

$y_i - d_i$ мәтінінің қарастырылған стильге қатыстығы.

Мәтін стилін тану үшін келесі мүмкіндіктер таңдалды:

1. Сөйлемдегі сөздердің орташа саны;
2. Мәтіндегі сөздің орташа ұзақтығы;
3. Мәтіндегі есімдіктер мен бөлшектерді білдіреді;
4. Белгілі бір префикстер мен жұрнақтардың көмегімен жасалған сөздердің орташа саны. Есепті шешудің негізі математикалық есептеулердің негізгі көлемі шоғырланған көпқабатты нейрондық желі (MNN) болды. Мәтін стилін тану үшін алға бағытталған нейрондық желі таңдалды.

Алғашқы үш мүмкіндік [7] алынған, өйткені олар өздерінің жоғары корреляциялық қасиеттерін көрсетті. Төртінші абзацта келтірілген

сөзжасамдық белгілер жанама сипатта болады. Әртүрлі стильдердің сөз формаларының әртүрлі саны бар деп болжанады.

Бұл мүмкіндіктерді есептеу үшін Python тілінде арнайы кітапханалар функциялары қолданылады. Алғашқы екі ерекшелік мәтінді жай сөйлемдер мен сөздерге бөлу, содан кейін сәйкес мәндерді санау арқылы есептеледі.

3-тармақтың мүмкіндіктері шектеулі сөздер жиынтығын қамтиды, сондықтан оларды іздеу үшін сіз барлық өкілдердің сөздігін жасап, осы сөздікте іздеуге болады. Ақырында, 4-тармақтың белгілерін іздеу үшін тұрақты тіркестерді қолданған тиімді. Қандай префикстер мен жұрнақтарды қолдану керектігін шешу үшін стильдің әрбір түрі үшін арнайы дайындалған мәтіндік корпустағы үлгілерді анықтау қажет [8].

Ұсынылған жұмыста алға қойылған мақсатқа жету үшін 1-суретте келтірілген модель жасалынған. Бұл модель 3 деңгейден тұрады: мәтін корпусын құру, мәтіндерді алдын ала өңдеу және тиімді алгоритмді анықтау. Мәтін корпусын құру және мәтіндерді алдын ала өңдеу процестері туралы біздің алдыңғы жұмыстарымызда толығырақ сипатталған [9-10]. Классификациялау әдістерінің ішінен K-жақын көршілер әдісі мен BERT таңдалған.

K – жақын көршілер әдісі (K-Near Neighbor (KNN) – бұл қарапайымдылығы мен тиімділігіне байланысты мәтінді алгоритмдеуде жиі қолданылатын әдістердің бірі. KNN – параметрлік емес жіктеу әдісі. Қысқаша айтқанда, KNN – бұл барлық қолданыстағы деректер объектілерін сақтайтын және ұқсастық өлшемі негізінде жаңа объектілерді жіктейтін қарапайым алгоритм. Мәтінді талдау саласында ол құжаттар мен k оқыту мәліметтері арасындағы ұқсастықты тексеру үшін қолданылады. Мақсаты – тесттік құжаттарының санатын анықтау. Мәтінді өндіруге арналған KNN-дің ең үлкен қосымшаларының бірі – бұл «тұжырымдамалық іздеу» (яғни мағыналық жағынан ұқсас құжаттарды табу), бұл бағдарламалық жасақтама құралы, бизнеске электрондық пошталарын, іскери хат-хабарларын, есептерін, байланыстарын және басқаларын табуға көмектеседі.

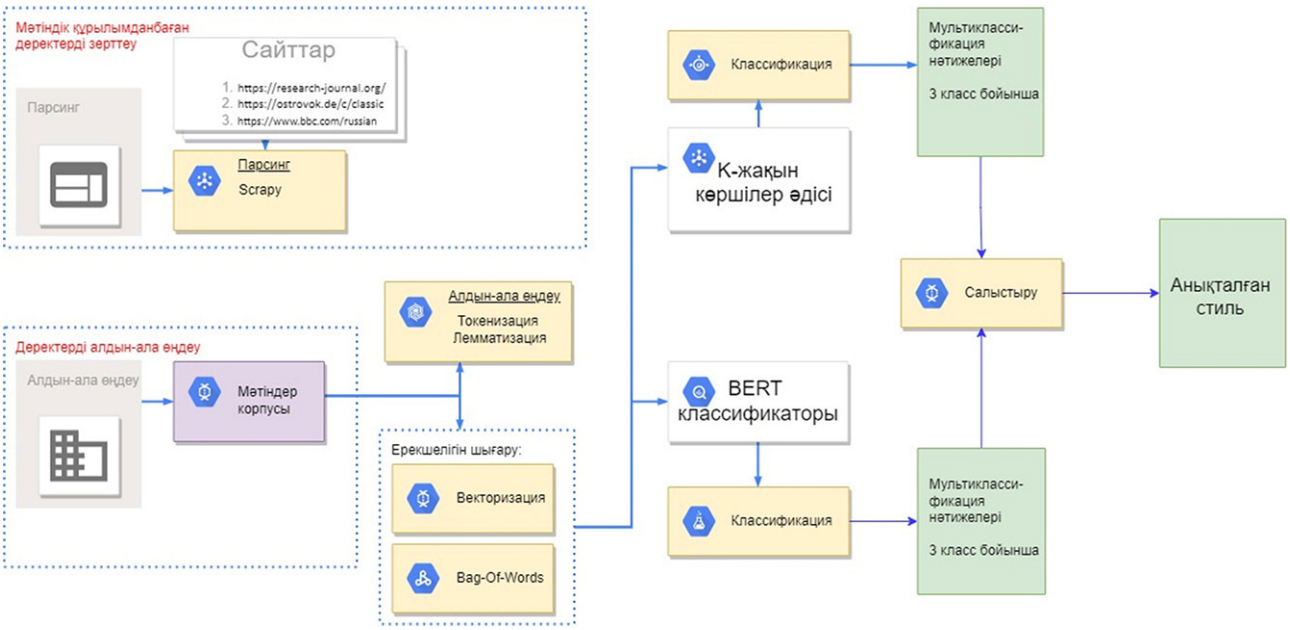
BERT (Bidirectional Encoder Representations from Transformers – Екі бағытты нейрондық желінің кодтаушысы) – Google іздеу жүйесіне пайдаланушылар енгізген сұраулардың контекстін талдауға мүмкіндік беретін әзірлеме. Алгоритм, атап айтқанда, табиғи тілді түсінуді жақсарту үшін жасалған. Екі бағытты BERT іздеуде сөз тіркестерін екі бағытта да талдайтынын білдіреді. Ол әрбір түйінді сөздің сол және оң жағындағы сөздерді ескереді. Осылайша, жүйе сөйлемдердің мағынасын түсінеді. Трансформатор есімдіктер мен көсемшелер сияқты сөздерді түсінуге көмектесетін тағайындау жүйесіне сілтеме жасайды.

Бұл әдістердің артықшылықтары мен кемшіліктері 1-кестеде келтірілген.

Талқылаулар және нәтижелер

Зерттеулерінің нәтижесінде құрылымданбаған деректертерді парсинг арқылы көшіру sience_literature_news.csv деректер жинағы құрылды. Мәтіндер сәйкесінше стиль түріне байланысты белгіленді: literature – көркем әдебиет стилі, scientific – ғылыми стиль, news – публицистикалық стиль.

Ресурстардың шектеулілігіне байланысты мәтіндердің жалпы саны 1243 болды (стильдері бойынша белгіленген мәтіндер). Әрі қарай барлық сан сәйкесінше 80%-дан 20%-ға дейінгі пайыз-



1-сурет – Мәтіндерді функционалдық стилі бойынша классификациялау моделі

дық қатынаста екі үлгіге (оқыту және тестілеу) бөлінді. Оқыту жиынтығы нейрондық желіні оқытуға арналған, ал тестілеу жиынтығы мәтін стилін тану сапасын тексеруге арналған.

Деректер жинағы бірінші әдіс К-жақын көршілер (KNN) әдісі арқылы классификацияланды. Жоғарыда айтылып кеткендей, 3 стиль таңдалғандықтан деректер 3 класқа топтастырылады. Модельдің архитектурасы төмендегідей анықталды:

```
from sklearn.neighbors import
KNeighborsClassifier
classifer = KNeighborsClassifier(n_neighbors=3)
classifer.fit(X_train, y_train)
y_pred = classifer.predict(X_test)
```

Нәтижесінде 96% дәлділік алынды (2-сурет).

Екінші таңдалған әдіс BERT моделін қолданатын классификатор болып табылады. BERT моделінің жоғалту және дәлдік қисықтары төмендегі 3-суретте келтірілген.

Эксперименттік жұмыстың нәтижелері таңдалған алгоритмдердің әрқайсысы мәтіндік құжаттардың функционалдық стилі бойынша классификациялауда айтарлықтай нәтижелер көрсеткенін көрсетті. 2-кестеде келтірілген модельдердің параметрлері бойынша талдау жүргізетін болсақ, белгіленген үш стильді анықтауда әр модель өз көрсеткіштерін берді. Яғни, мәтіннің көркем әдебиет стилі жататындығын KNN әдісі, ғылыми және публицистикалық стильдерді BERT моделі жақсы анықтайтындығы көрсетілген. Бұл BERT – Google іздеу алгоритмінің соңғы қосымшасы және соңғы бес жылдағы осы саладағы ең кең

1-кесте – Мәтінді жіктеудің әртүрлі әдістеріне шолу

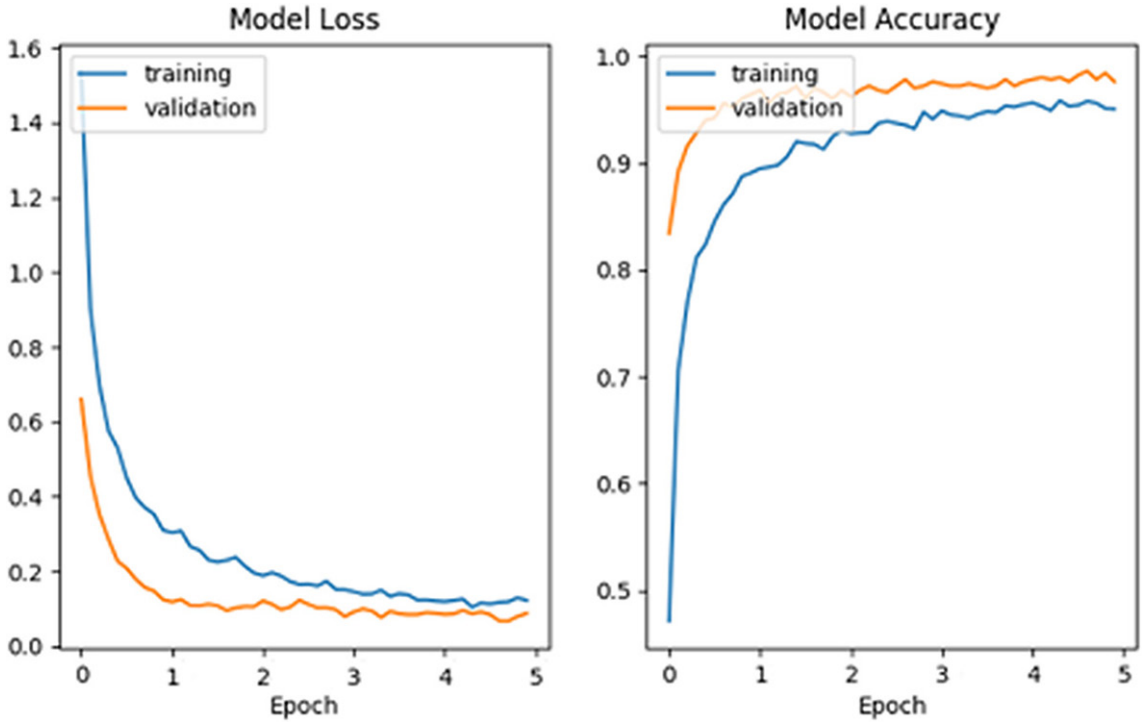
Әдіс атауы	Қасиеттері/Параметрлер		
	Дәлдік	Оқыту уақытының тездігі	Сызықтық
Логистикалық регрессия	-	+	+
SVM	-	+	+
К-жақын көршілер әдісі	+	+	+
Шешімдер ағашы	+	-	-
BERT	+	+	+
Аңғал Байес	-	-	+

```
Confusion Matrix:
[[11  0  0]
 [ 0 12  1]
 [ 0  0  6]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        11
           1       1.00      0.92      0.96        13
           2       0.86      1.00      0.92         6

    accuracy                           0.97        30
   macro avg       0.95      0.97      0.96        30
weighted avg       0.97      0.97      0.97        30

Accuracy: 0.9666666666666667
```

2-сурет – К-жақын көршілер әдісіне негізделген модель классификацисының нәтижесі

2-кесте – К-жақын көршілер (KNN) және BERT модельдерінің нәтижелерін салыстыру

Стиль түрлері	Модель	Дәлдігі	Толықтық	F-өлшемі
Көркем әдебиет стилі	KNN	1.00	1.00	1.00
	BERT	1.00	0,98	0,99
Ғылыми стиль	KNN	1.00	0.92	0,96
	BERT	1.00	1.00	1.00
Публицистикалық стиль	KNN	0.86	1.00	0.92
	BERT	1.00	0.89	0,94



3-сурет – BERT моделінің жоғалту және дәлдік қисықтары

қолданыстағы модель екендігін айқын дәлелдеді. Сонымен қатар, ең жақсы классификаторды таңдау үшін «сынаулар мен қателер» санын осы зерттеуде берілген ақпарат негізінде азайтуға болады.

Қорытынды

Мәтіндер стилін анықтауда қолданылатын мәтіндік деректер көздері қарастырылды. Мәтіндер корпусы толық өңделіп, қажетті әдістерді қолдана отырып эксперименттер жасалды, программалық реализация орындалды. Мәтіндер K-жақын көршілер және BERT классификация модельдері қолданылды. Модельдердің түрлері белгілі бір функционалдық стильге тиімдірек екені көрсетілді. Модельдер оқыту параметрлері бойынша салыстырылды.

ӘДЕБИЕТТЕР ТІЗІМІ

1. Бенджамин Б., Ребекка Б., Тони О. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – СПб: Издательский дом «Питер», 2018. – 368 с.
2. Цитульский А.М., Иванников А.В., Рогов И.С. NLP-обработка естественных языков // StudNet. – 2020. – Т. 3. – № 6. – С. 467-475.
3. Pal S., Ghosh S., Nag A. Sentiment analysis in the light of LSTM recurrent neural networks // International Journal of Synthetic Emotions (IJSE). – 2018. – Т. 9. – No. 1. – Pp. 33-39.
4. Thangaraj M., Sivakami M. Text classification techniques: a literature review // Interdisciplinary Journal of Information, Knowledge, and Management. – 2018. – Т. 13. – P. 117.
5. Ferreira C.H.P., De Franca F.O., Medeiros D.R. Combining multiple views from a distance based feature extraction for text classification // 2018 IEEE Congress on Evolutionary Computation (CEC). – IEEE, 2018. – Pp. 1-8.
6. Sari W.K., Rini D.P., Malik R.F. Text Classification Using Long Short-Term Memory with GloVe // Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI). – 2019. – Т. 5. – No. 2. – Pp. 85-100.
7. Gurney K. An introduction to neural networks. – CRC press, 2018.
8. Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016.
9. Кайбасова Д.Ж., Маханова Б., Сулеймен А.Е. Қазақ тіліндегі мәтін стилін анықтауда табиғи тілді өңдеу (NLP) әдістері // Труды университета. № 2 (83), 2021, С. 172-176, DOI 10.52209/1609-1825_2021_2_172
10. D. Kaibassova, L. La, A. Smagulova, L. Lisitsyna, A. Shikov, M. Nurtay, «Methods and algorithms of analyzing syllabuses for educational programs forming intellectual system», Journal of Theoretical and Applied Information Technology, 15th March 2020, Vol. 98, No. 05, pp. 876-888.

Оптимальная модель классификации текстов по функциональному стилю

[1]***КАЙБАСОВА Динара Женисбековна,*** *PhD, и.о. доцента, dindgin@mail.ru,*
[1]***ОЛЕЙНИКОВА Алла Васильевна,*** *магистр, старший преподаватель, alla_ole@mail.ru,*
[1]***МАХАНОВА Балжан Мақсатқызы,*** *магистрант, balzhan.makhanova@gmail.com,*
[1]НАО «Карагандинский технический университет имени Абылкаса Сагинова», Казахстан, Караганда, пр. Н. Назарбаева, 56,

**автор-корреспондент.*

Аннотация. *Сегодня обработка естественного языка развивается благодаря значительному улучшению доступа к данным и увеличению вычислительной мощности, особенно в таких областях, как здравоохранение, средства массовой информации, финансы и человеческие ресурсы. Целью данной статьи является изучение методов интеллектуального анализа текстов при определении стиля текстов и создание оптимальной модели. Рассмотрены источники текстовых данных, используемые для определения стиля текстов. Полностью обрабатывается тело текста и проводятся эксперименты с использованием необходимых методов. Для классификации текстов использовались модели K-neighbours и BERT.*

Ключевые слова: *анализ текста, стиль текста, функциональный стиль, классификация, K-ближайшие соседи, BERT, естественные языки, интеллектуальный анализ текста, литературный стиль, научный стиль, публицистический стиль.*

Optimal Model for Classifying Texts by Functional Style

[1]***KAIBASSOVA Dinara,*** *PhD, Acting Associate Professor, dindgin@mail.ru,*
[1]***OLEINIKOVA Alla,*** *Master, Senior Lecturer, alla_ole@mail.ru,*
[1]***MAKHANOVA Balzhan,*** *Master Student, balzhan.makhanova@gmail.com,*
[1]NPJSC «Abylkas Saginov Karaganda Technical University», Kazakhstan, Karaganda, N. Nazarbayev Avenue, 56,

**corresponding author.*

Abstract. *Nowadays natural language processing is evolving due to significant improvements in data access and increased computing power, especially in areas such as healthcare, media, finance and human resources. The purpose of this article is to study the methods of intellectual analysis of texts in determining the style of texts and to create an optimal model. The sources of text data used to determine the style of texts are considered. The body of the text is fully processed and experiments are carried out using the necessary methods. K-neighbors and BERT models were used for text classification.*

Keywords: *analysis, text style, functional style, classification, K-nearest neighbors, BERT, natural languages, text mining, literary style, scientific style, journalistic style.*

REFERENCES

1. Benjamin B., Rebecca B., Tony O. Applied Text Analysis with Python. – Saint Petersburg: Publishing House «Piter», 2018. – 368 p.
2. Zitulski A.M., Ivannikov A.V., Rogov I.C. NLP-obrabotka estestvennih yazikov // StudNet. – 2020. – T. 3. – No. 6. – Pp. 467-475.
3. Pal S., Ghosh S., Nag A. Sentiment analysis in the light of LSTM recurrent neural networks // International Journal of Synthetic Emotions (IJSE). – 2018. – T. 9. – No. 1. – Pp. 33-39.
4. Thangaraj M., Sivakami M. Text classification techniques: a literature review // Interdisciplinary Journal of Information, Knowledge, and Management. – 2018. – T. 13. – P. 117.
5. Ferreira C.H.P., De Franca F.O., Medeiros D.R. Combining multiple views from a distance based feature extraction for text classification //2018 IEEE Congress on Evolutionary Computation (CEC). – IEEE, 2018. – Pp. 1-8.
6. Sari W.K., Rini D.P., Malik R.F. Text Classification Using Long Short-Term Memory with GloVe // Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI). – 2019. – T. 5. – No. 2. – Pp. 85-100.
7. Gurney K. An introduction to neural networks. – CRC press, 2018.
8. Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016.
9. Kaibassova D., Makhanova B., Syleimen A. Kazakh titltndegi matin stylin anyktauda tabigi tilde ondeu (NLP) adistery // Trudy University. No. 2 (83), 2021, pp. 172-176, DOI 10.52209/1609-1825_2021_2_172
10. D. Kaibassova, L. La, A. Smagulova, L. Lisitsyna, A. Shikov, M. Nurtay, «Methods and algorithms of analyzing syllabuses for educational programs forming intellectual system», Journal of Theoretical and Applied Information Technology, 15th March 2020, Vol. 98, No. 05, pp. 876-888.